WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of Twitter Sentiments About the Russian-Ukraine War Using Naive Bayes Based on Particle Swarm Optimization

Setefensius Sasi [1], Ema Utami[2], Eko Pramono[3]

[1]Magister of Informatics Engineering, *AMIKOM University*Yogyakarta, Indonesia
[2]Magister of Informatics Engineering, AMIKOM University Yogyakarta, Indonesia
[3]Magister of Informatics Engineering, AMIKOM University Yogyakarta, Indonesia
DOI: https://doi.org/10.55248/gengpi.2023.4138

Abstract—

The Russia-Ukraine crisis has not found a solution until now, at least until November 2022. Many Indonesians have expressed their opinions on this matter via social media Twitter. This study uses the Naïve Bayes (NB) Algorithm based on Particle Swarm Optimization (PSO) to analyze opinions regarding the Russia-Ukraine war. Data was taken from Twitter using the keywords "Russia Ukraine", "Russia vs Ukraine", "Russia-Ukraine", and "Perang Russia". The number of data sets taken is 5000 data. The results showed that the accuracy of the Naïve Bayes (NB) algorithm without Particle Swarm Optimization (PSO) was 67.72%, the precision value was 58.33%, the recall value was 79.75%, and the error rate of 57.14% while Naïve Bayes (NB) with Particle Swarm Optimization (PSO) the accuracy obtained is 73.48%, the precision value is 65.62%, the recall value is 76.36% and error rate of 50.36%. Thus it can be said that the Particle Swarm Optimization (PSO) applied to the Naïve Bayes algorithm for objects in research can increase the accuracy of results.

Keywords— Russia Ukraine War; Sentiment analysis; Naïve Bayes; Particle Swarm Optimization

I. Introduction

The Russia-Ukraine conflict can be seen as an internal problem for the two countries. However, the existence of Ukraine as an entity that is quite important on the border with Russia, does have not only strategic value for Russia but also for Western European countries, and even the world[1]. So that the Ukrainian crisis also invites the interests of many parties, including the US which is a member of NATO along with European Union countries[2].

The Russia-Ukraine crisis, which has not found a solution until now, at least until November 2022, has raised international concern considering that war or open conflict has the potential to occur if the conflicting parties cannot control themselves anymore[3]. The end of this episode of the Russia-Ukraine crisis cannot be predicted clearly, because the outcome still depends on the efforts of the leaders, the agreement of the two camps, and the defense ego of each country which could explode resulting in war [4].

Therefore, the international community must continue to seek peaceful solutions through dialogue and diplomacy so that the threat of another world war does not arise. As a member of the international community, Indonesia with its free and proactive foreign policy must participate in finding the best solution to resolve the conflict between Russia and Ukraine [5]. This is because the ongoing conflict will ultimately harm many parties, not only in Europe but also in other regions. The disturbing relationship can damage cooperation in the economic and trade fields related to the economic interests of the people in each country[6].

The war between Russia and Ukraine has had a range of adverse effects that threaten the economy in all sectors, especially in terms of global trade [7]. The two countries are known as the largest suppliers of various major commodities such as wheat, fertilizer, natural gas, and oil. If these goods are required by international consumers. As a result of the blockade imposed on Ukraine by Russia, various types of goods cannot be distributed and import and export activities are hampered [8].

In this regard, this article aims to analyze the discourse of the Indonesian people about the Russia-Ukraine crisis, especially their opinions on social media Twitter. Many people expresstheir opinions and expressions through technology such as social media [9]. And one of the social media that is widely used by the public in expressing opinions and expressions is Twitter. This is important to analyze considering that Indonesia is the host of the G20, so Indonesian public discourse must be known. Thus, the Indonesian government can make appropriate decisions and policies for the sake of national security and the smooth functioning of the G20 activities is beneficial [8].

By utilizing data from Twitter social media, an analysis of the opinions and opinions of the Indonesian people regarding the Russia-Ukraine crisis can be carried out through sentiment analysis by classifying opinions and opinions into 2 classes, namely negative and positive [10]. The discourse developed by the Indonesian people on social media Twitter is interesting and important to analyze because posting opinions and opinions of politicians on Twitter can be used as an instrument to measure the direction of national policy.

Sentiment analysis is a method of measuring and analyzing specific cases and items, and it is possible to make sentiment analysis conclusions and judgments based on text such as sentences and documents. Some methods or algorithms that can be used for sentiment analysis are Naive Bayes, Support Vector Machines, etc. Naive Bayes is a classification method derived from Bayes' theorem. Developed by British scientist Thomas Bayes, this method of classification using probability and statistics, known as Bayes' theorem, predicts future probabilities based on past experience [11].

Data collection in this study was carried out by retrieving data sets from Twitter by utilizing the API facility [12]. The data set was retrieved using several keywords related to the Russia-Ukraine war including “rusiaukraina”, “rusia vs ukraina”, “rusia-ukraina”, “perangrusia”, and "perangrusia". The Naive Bayes algorithm was chosen because previous research explained that Bayes is an algorithm that is quite easy to implement but has a fairly good level of performance in sentiment analysis.

II. Literature Review

A previous study by EmaUtami et al. (2021), titled Analysis of Post-Covid-19 Online College Twitter Sentiment Using Support Vector Machines and Naive Bayes Algorithms, compared Naive Bayes and Support Vector Machines (SVMs) to find the best Find way to classify. The performance results obtained with Bayes are 81.20% accuracy, 9.00 seconds in time, 79.60% recall, and 79.40% accuracy, and the SVM algorithm has an accuracy score of 85%, and 31.60 in time. seconds. 84% and 83.60% accuracy, performance results were obtained for his 1st iteration of Naive Bayes and the 423rd iteration of his SVM algorithm [13].

A study by FrizkaFitriana et al. (2021) Opinion Sentiment Analysis of His Covid-19 Vaccine on Twitter Social Media Using Support Vector Machines and Naive Bayes. This study compared Naive Bayes and Support Vector Machine (SVM) algorithms to conduct sentiment analysis on Covid-19 vaccines. Based on our research results, we find that the SVM algorithm performs well in terms of accuracy, precision, and recall with a score of 90.47%. 90.23%, 90.78%, Naive Bayes' performance is 88.64%, 87.32%, 88.13%, with a difference of 1.83 - accuracy, 2.91% accuracy, and 2.65% recall algorithm Naive Bayes has a time rate of 11 seconds A value of 8.1 seconds compared to the SVM with Sentiment analysis results were 8.76% Neutral, 42.92% Negative, 48.32% Positive for Naive Bayes and 10.56% Neutral, 41.28% Negative, 48.16% Positive for SVM. [10].

A study by Yuris et al. (2020) describes Support Vector Machine (SVM) and Naive Bayes (NB) algorithms based on Particle Swarm Optimization (PSO) in Sentiment Analysis for the elimination of the national exams. For this study, we started by collecting Twitter sentiment data. The data were processed using four different methods: SVM without PSO, SVM with PSO, NB without PSO, and BN with PSO. SVM with PSO has the highest accuracy value compared to the other three methods with accuracy data of 92.92, the other accuracy values are 94.81% for SVM without PSO and NB without PSO We can conclude that 85.9% and 86.92%. NB with PSO [14].

Ratino et al. (2020) used Support Vector Machines (SVM) and Naive Bayes (NB) algorithms to analyze public opinion conveyed by Instagram social media comments on COVID-19 information. Naive Bayes gives 78.02% accuracy and 0.714 AUC, while support vector machines give 80.23% accuracy and 0.904 AUC. There is an accuracy difference of 2.21%. After optimization with the Particle Swarm Optimization operator, the Naive Bayes (PSO) algorithm returned 79.07% accuracy and an AUC of 0.729, while the Support Vector Machine (PSO) algorithm returned 81.16% accuracy and achieved an AUC of 0.903. There is an accuracy difference of 2.09%. Algorithm test results can always yield higher accuracy whether it is a PSO-based support vector machine or not [9].

III. Research Methods

A. Model

In this study, we go through several steps, including taking tweet data from the Twitter site, preprocessing it, and then performing a manual labeling process before entering the training process. After the classification model is formed in the training process, the classification process is performed[15]. The research stages are shown in Figure 1.

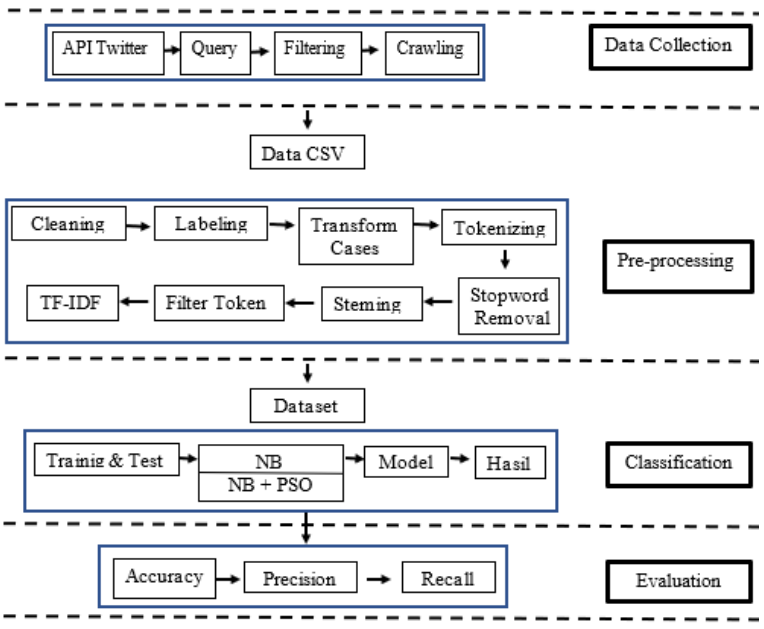

Fig. 1 Research Model

B. Data Collection Techniques

Data collection was carried out by crawling tweet data using the Application Program Interface (API) facility provided by Twitter. Twitter Search API has the disadvantage that it can only search for tweets published in the last seven days. The collected data is then stored in a .csv file. Then the tweets used as training data are labeled positive and negative[16].

C. Data Processing

- The next stage is pretreatment. Preprocessing is a method performed before running the data mining process to generate a more superficial meaning. The pretreatment process is divided into 9 steps [14]:
- Cleaning to remove unnecessary variables such as URLs, symbols, etc
- Labeling to label positive, negative, and neutral. Labeling will be done manually
- Transform Case changes all uppercase or capital letters in the data to lowercase letters so that there is uniformity
- Tokenizing to break sentences into several parts or words
- Stopword removal to remove unnecessary words because if the word is removed it will not change the information contained in the sentence, for example, conjunctions yang, akan, di, pada, etc.
- Stemming to remove the affixes contained in the word with the aim that the word returns to the base word, for example, the word "menulis" is changed to the word "tulis"
- Filter Tokens omit words of a certain letter length. For example, a minimum of 4 characters and a maximum of 25 characters. This means that words that are only 4 characters long and more than 25 characters will be removed
- Term Frequency-Inverse Documents Frequency (TF-IDF) to Give weight to text
- Split data to divide the dataset into several parts.

D. Data Modeling

In this phase, the preprocessed data set is classified. Determine accuracy, precision, and recall by testing with the Rapidminer tool. The modeling uses a naive Bayes (NB) algorithm based on particle swarm optimization (PSO). A Naive Bayes classifier is a supervised classifier because it has a supervisor (a human manual classification based on the data used in training) as a teacher during the learning process[11].

A naive Bayes classifier is a classification method based on Bayes' theorem. This classifier assumes that the presence of features within a class is independent of other features [15]. Equation (1) is a Bayesian formula.

$$P(Y|X) = \frac{P(X|y)P(Y)}{P(X)} \quad (1)$$

Where P(X|Y) is the probability of occurring X if it is known Y. P(Y|X) is the chance of occurring Y if it is known X. P(X) is the probability of occurring X and P(Y) is the probability of occurring Y.

E. Evaluation

We also use a confusion matrix to measure the accuracy of the created classifier model. Confusion matrices are essential tools in visualization techniques used in machine learning, which usually contain two or more categories. A confusion matrix is a table that provides information in the form of a comparison of classification results derived from "predictions". with actual classification results. The Confusion Matrix table shows the number of correctly classified test data and the number of incorrectly classified test data. The evaluation phase is run to determine the accuracy of the modeling applied to the training data. Then compare the results for two different datasets by applying the confusion matrix and calculating the precision[15].

IV. Results and Discussion

A. Retrieval of Dataset

The process of data retrieval (crawling) from Twitter using the RapidMiner Application with the Twitter search operator facility with the query "Russia Ukraine", "Russia vs. Ukraine", "Russia-Ukraine", "Perang Russia". Then the data will be saved in CSV format using the write CSV operator. The number of data sets taken is 5000 data.

The parameters used in RapidMiner are the twitter-connection parameters which function to connect RapidMiner with Twitter through the Application Program Interface (API) facility provided by Twitter. Then the Search Twitter parameter functions to perform queries. The attribute select parameter is used to filter the attributes needed from Twitter data, namely text data only. Meanwhile, the write CSV parameter aims to store crawled data in CSV format.

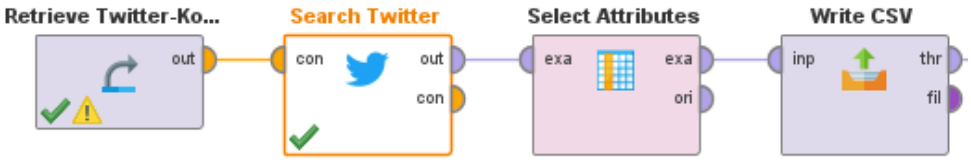

Fig. 2 Crawling data from Twitter

B. Data Processing

After the data has been successfully saved into excel format, then data cleansing is performed using several operators in RapidMiner. After cleaning, the data set that is ready to be labeled is 1260 data. 882 data is labeled manually which is then called training data. The remaining 378 data are test data. Figure 2 shows the data processing sequence.

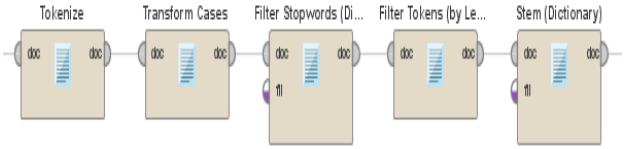

Fig. 3 Data Processing

After processing the data, the dataset will be divided into 70% (882 data) training data and 30% (378 data) test data as shown in Figure 3

Fig. 4 Modeling Data Distribution

C. Modeling

Data The data used is to classify positive, and negative labels on tweets related to the Russia-Ukraine war. The algorithms used are Naïve Bayes and Naïve Bayes using Particle Swarm Optimization (PSO) to get the best accuracy value. Figure 4 shows the Naïve Bayes data modeling based on Particle Swarm Optimization (PSO).

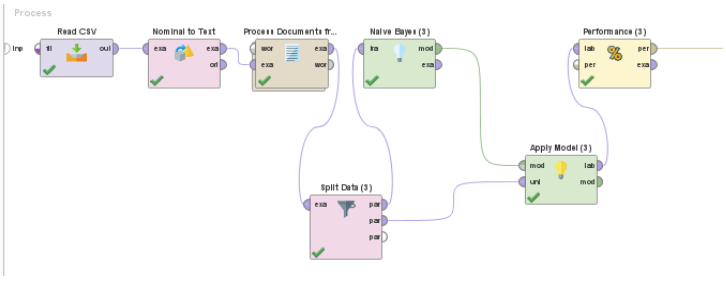

Fig. 5 ModelingData Naïve Bayes

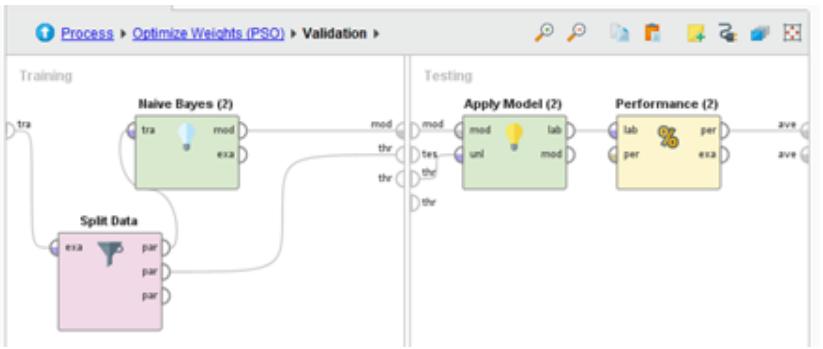

Fig. 6 ModelingData Naïve Bayes with PSO

The Optimation was performed by comparing two results of modeling a Naive Bayes (NB) algorithm without Particle Swarm Optimization (PSO) and a Naive Bayes (NB) algorithm with Particle Swarm Optimization (PSO). increase. The purpose of the optimation is to determine the practical value of the model successfully created in the previous step. From the results of model testing using the Naive Bayes (NB) algorithm without Particle Swarm Optimization (PSO), we can generate the accuracy values (confusion matrix) shown in Figure 5.

accuracy: 67.72%

	true Positif	true Negatif	class precision
pred. Positif	130	32	80.25%
pred. Negatif	90	126	58.33%
class recall	59.09%	79.75%	

Fig. 7 Accuracy Naïve Bayes

For model testing using the Naïve Bayes algorithm (NB) with Particle Swarm Optimization (PSO) can produce an accuracy value (confusion matrix) which can be seen in Figure 6

accuracy: 73.48%

	true Positif	true Negatif	class precision
pred. Positif	55	13	80.88%
pred. Negatif	22	42	65.62%
class recall	71.43%	76.36%	

Fig. 8 Accuracy of Naïve Bayes with Particle Swarm Optimization

The following is a comparison of the accuracy, precision, and recall values of model testing using the algorithm Naïve Bayes (NB) and Naïve Bayes (NB) with Particle Swarm Optimization (PSO).

Table I Table of Comparison of Confusion matrix Value

Model	Accuracy	Precision	Recall
Bayes (NB)	67.72%	58.33%	79.75%
(NB) + PSO	73.48%	65.62%	76.36%

D. Evaluation

The evaluation uses a confusion matrix, that is, a true positive rate, a true negative rate, a false positive rate, and a false negative rate as indices. The true positive rate is the percentage of positive classes that are successfully classified as positive classes, and the true negative rate is the percentage of negative classes that are successfully classified as negative classes. The same is true for the false rate, which is the opposite of the true rate.

Table II Confusion Matrix Naïve bayes

Accuracy: 67.72%			
	True Positif	**True Negatif**	**Class Precision**
Pred. Positif	130	32	80.25%
Pred. Negatif	90	126	58.33%
Class Recall	59.09%	79.75	

The accuracy results obtained are 67.72% of the 220 positive tweet data and 158 negative tweets about the Russia-Ukraine war. Positive tweet data that corresponds to a positive prediction (TP) is 130 data. While the negative data that goes into the positive prediction (TN) is 32 data. The positive tweet data that goes into the negative prediction (FP) is 90 data and the negative tweet data that fits into the negative prediction (FN) is 126 data. So 67.72% is the result of (130+126) divided by 378 then multiplied by 100%. The error rate of 57.14%.

Table III Confusion Matrix Naïve bayes with pso

Accuracy: 73.48%			
	True Positif	**True Negatif**	**Class Precision**
Pred. Positif	55	13	80,88%
Pred. Negatif	22	47	65,62%
Class Recall	71,43%	76,36	

The accuracy results obtained are 73,48% of the 77 positive tweet data and 60 negative tweets about the Russia-Ukraine war. Positive tweet data that corresponds to a positive prediction (TP) is 55 data. While the negative data that goes into the positive prediction (TN) is 13 data. The positive tweet data that goes into the negative prediction (FP) is 22 data and the negative tweet data that fits into the negative prediction (FN) is 47 data. So 73.48% is the result of (55+47) divided by 137 then multiplied by 100%. The error rate of 50.36%.

Conclusions and Suggestions

Based on the results of tests performed on the Russian-Ukrainian war dataset using the Naive Bayes (NB) algorithm without using Particle Swarm Optimization (PSO), the obtained accuracy is 67.72%. , with an precision value of 58.33%, a recall value of 79.75%, and error rate of 57.14%. Naive Bayes (NB) using Particle Swarm Optimization (PSO) had an accuracy of 73.48%, an precision value of 65.62%, a recall of 76.36%, and error rate of 50.36%. Thus, it can be said that the particle swarm optimization (PSO) applied to the studied Naive Bayes algorithm can improve the accuracy of the results.

References

[1]. A. Y. Prasetyo, “Apa yang Dicari Vladimir Putin di Ukraina?,” kompas.id, Feb. 12, 2022.

[2]. S. Hidriyah, “KAJIAN SINGKAT TERHADAP ISU AKTUAL DAN STRATEGIS ESKALASI KETEGANGAN RUSIA-UKRAINA,” 2022.

[3]. L. A. Saputra, “TensiMeningkat, Sejumlah Negara Minta WarganyaTinggalkanUkraina,” kompas.id, Feb. 13, 2022.

[4]. D. M. (Ed.)Malitz and S. (Ed.) Sriyai, “Civil Societies’ Reactions of East-and Southeast Asian Countries to the Russian War on Ukraine: Preliminary Observations,” 2022. [Online]. Available: https://www.uni-due.de/in-east/news/green_series.php

[5]. B. Rafie, “PerangBisa Meletus Kapan Saja, IniAsalMulaPertikaianRusia dan Ukraina,” internasional.kontan.co.id, Feb. 12, 2022.

[6]. C. R. Bakrie, M. O. Delanova, and Y. M. Yani, “PENGARUH PERANG RUSIA DAN UKRAINA TERHADAP PEREKONOMIAN NEGARA KAWASAN ASIA TENGGARA,” 2022. [Online]. Available: https://sridianti.com/konferensi-pbb-tentang-perdagangan-dan-pembangunan-

[7]. D. Tiara Kusuma, S. Az-Zahra, M. Huda Apriyadi, and B. Siregar, "REAKSI PASAR MODAL INDONESIA TERHADAP PERISTIWA PERANG RUSIA DAN UKRAINA," JurnalIlmiah MEA (Manajemen, Ekonomi, &Akuntansi), vol. 6, no. 3, p. 2022, 2022.

[8]. G. F. I. Hutabarat, "ArahKebijakanLuar Negeri Indonesia PascaPerangRusia-UkrainaBerdasarkanPerspektif National Interest," Jurnal Al Azhar Indonesia Seri IlmuSosial, vol. 3, no. 3, p. 154, Oct. 2022, doi: 10.36722/jaiss.v3i3.1323.

[9]. Ratino, N. Hafidz, S. Anggraeni, and W. Gata, "SentimenAnalisisInformasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes," 2020.

[10]. F. Fitriana, E. Utami, and H. al Fatta, "AnalisisSentimenOpiniTerhadapVaksinCovid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," JurnalKomtika (Komputasi dan Informatika), vol. 5, no. 1, pp. 19–25, Jul. 2021, doi: 10.31603/komtika.v5i1.5185.

[11]. Hernawati and W. Gata, "SENTIMEN ANALISIS OPERASI TANGKAP TANGAN KPK MENURUT MASYARAKAT MENGGUNAKAN ALGORITMA SUPPORT VECHTOR MACHINE, NAÏVE BAYES, BERBASIS PARTICLE SWARM OPTIMIZITION," Faktor Exacta, vol. 12, no. 3, p. 230, Nov. 2019, doi: 10.30998/faktorexacta.v12i3.4992.

[12]. G. A. Buntoro, "Sentiments analysis for prediction the governor of east java 2018 in twitter," International Journal Artificial Intelligent and Informatics, vol. 1, no. 2, pp. 82–90, Dec. 2018, doi: 10.33292/ijarlit.v1i2.20.

[13]. H. Setiawan, E. Utami, and S. Sudarmawan, "AnalisisSentimen Twitter Kuliah Online Pasca Covid-19 MenggunakanAlgoritma Support Vector Machine dan Naive Bayes," JurnalKomtika (Komputasi dan Informatika), vol. 5, no. 1, pp. 43–51, Jul. 2021, doi: 10.31603/komtika.v5i1.5189.

[14]. Y. Alkhalifi, W. Gata, A. Prasetyo, and I. Budiawan, "AnalisisSentimenPenghapusanUjian Nasional pada Twitter Menggunakan Support Vector Machine dan Naïve Bayes berbasis Particle Swarm Optimization," JurnalCoreIT, vol. 6, no. 2, 2020.

[15]. S. Rahmawati and M. Habibi, "Public Sentiments Analysis about Indonesian Social Insurance Administration Organization on Twitter," IJID (International Journal on Informatics for Development), vol. 9, no. 2, pp. 87–93, Dec. 2020, doi: 10.14421/ijid.2020.09205.

[16]. M. S. Mubarok, A. Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," in AIP Conference Proceedings, Aug. 2017, vol. 1867. doi: 10.1063/1.4994463